U.S. Fish and Wildlife Service
Columbia River Fisheries Program Office

Factors Influencing Passive Integrated Transponder (PIT) Detection Efficiency in Tryon Creek, 2015 Annual Report

Brian Davis, Jen Poirier, Steve Haeseker and Courtney Newlon

U.S. Fish and Wildlife Service
Columbia River Fisheries Program Office
Vancouver, WA 98683

On the cover: *Image of Tryon Creek looking upstream, featuring a beaver dam located between Antennas #1 and #2.*

Disclaimers

The correct citation for this report is:

Davis, M.B., J. Poirier, S. Haeseker, and C. Newlon. 2016. Factors Influencing Passive Integrated Transponder (PIT) Antenna Detection Efficiency in Tryon Creek, 2015 Annual Report. U.S. Fish and Wildlife Service, Columbia River Fisheries Program Office, Vancouver, WA. 18 pp.

Factors Influencing Passive Integrated Transponder (PIT) Detection Efficiency in Tryon Creek
2015 ANNUAL REPORT

M. Brian Davis, Jen Poirier, Steve L. Haeseker and Courtney Newlon

U.S. Fish and Wildlife Service
Columbia River Fishery Program Office
1211 SE Cardinal Court, Suite 100
Vancouver, WA 98683

Abstract - Passive Integrated Transponder (PIT) tag detection systems are useful tools that aid our understanding of the movement, survival and abundance of fish populations. Increasing antenna detection efficiency can improve fish population estimates, providing more accurate information to biologists and managers. In 2015 we monitored biological, electrical, environmental, and physical factors in Tryon Creek, a small urban stream in Portland Oregon, in an effort to identify factors that influence detection probability and to assess the relative importance of those factors. Several classification models were evaluated, using k-fold cross-validation to fit each model and to compare the models' prediction success rates. Flow was identified as the most influential contributor to PIT antenna detection probabilities; as flow (cfs) increased per unit, the expected odds of detection decreased by 2.1 percent. Time of detection, tagging date, and some PIT transceiver settings (i.e., phase and capacitance) were also found to be influential, but less so. Only two arrays were operational during the six month span of tagging efforts, and we did not monitor and/or analyze some factors we intended to due to static environmental conditions and improper transceiver settings. We plan to continue monitoring in 2016 with two more arrays in place, past mistakes remedied and an added comparison between pass-over and pass-through configured PIT arrays.

Page intentionally left blank

Table of Contents

List of Tables

List of Figures

Introduction

Passive Integrated Transponder (PIT) tags and their associated detection systems have emerged as a valuable technology for monitoring the movement, survival, and abundance of fish populations. Because the precision and accuracy of demographic rate estimates in mark-recapture studies is related to the number of tagged fish that are subsequently detected, it is important to maximize the efficiency of PIT detection arrays and understand the factors that influence detection efficiency. Ideally, antenna efficiency should be high and harbor low bias. Although physical, environmental, electrical, and biological factors likely influence the realized detection efficiency, the relative importance of these factors is rarely assessed in stream and small river studies. In large rivers where PIT arrays have been installed in hydropower dams, Plumb et al. (2012) and McCann et al. (2015) have shown that detection efficiency can vary with flow, hydropower operations, and by species. However, in stream and small river studies, where most PIT arrays are installed, the factors that influence detection efficiency have not been thoroughly investigated.

In the spring of 2015, we developed a study plan with goals to measure and analyze a number of metrics in order to gain knowledge of PIT antenna efficiency, the factors affecting efficiency and the relative importance of these factors. Tryon Creek was selected as our preliminary study site due to ongoing intensive PIT tagging efforts, two fully operational PIT antennas, readily available discharge data, and its close proximity to the Columbia River Fisheries Program Office (CRFPO). Three objectives were proposed for this study site to be initiated in 2015:

1. Collect data on the biological, electrical, environmental and physical factors that may influence detection probability.
2. Identify important variables and attempt to answer these questions:
   - Which antenna performed best and why?
   - Which factors have the greatest influence on detection probability and why?
   - How can we effectively improve antenna efficiency?
   - Can we use other means besides detection history to predict antenna efficiency?
3. Increase detectability and accuracy of antenna efficiency estimates.

Study Site

Tryon Creek is a 7.81 km second-order urban stream located in southwest Portland, Oregon that flows through residential neighborhoods and Tryon Creek State Natural Area before joining the Willamette River at rkm 32 (see Figure 1). The CRFPO is currently using PIT technology to assess species community, abundance and temporal use of fish in the lower 0.3km stretch of recently enhanced confluence habitat immediately downstream from the Highway 43 culvert (see Silver et al. 2014 for project details).

The Tryon Creek PIT in-stream detection system is located about 140 meters upstream of the Tryon Creek's confluence with the Willamette River. The site originally consisted of two pass through FDX antennas powered by a single FS1001M multiplexing transceiver (MUX). On 16 June 2015, two additional antennas were built and installed upstream of the existing antennas (totaling 4). Three of the four antennas were configured in a pass-through orientation (Antennas #1, #3 and #4) with Antenna #2 being configured as a pass-over (see Figure 2). All pass-through antennas spanned the wetted width during normal flows, were about 1 meter in height and read 12mm PIT tags 100% through the antenna's center. Antenna #2 (pass-over) spanned the wetted width, but could only read 12mm tags 18cm above the top of the antenna. PTAGIS naming convention requires antennas to be numbered sequentially with the most upstream antenna be being named #1, which can result in the renaming of antennas as more are added. To help alleviate confusion, we have given each antenna location a unique proper name.

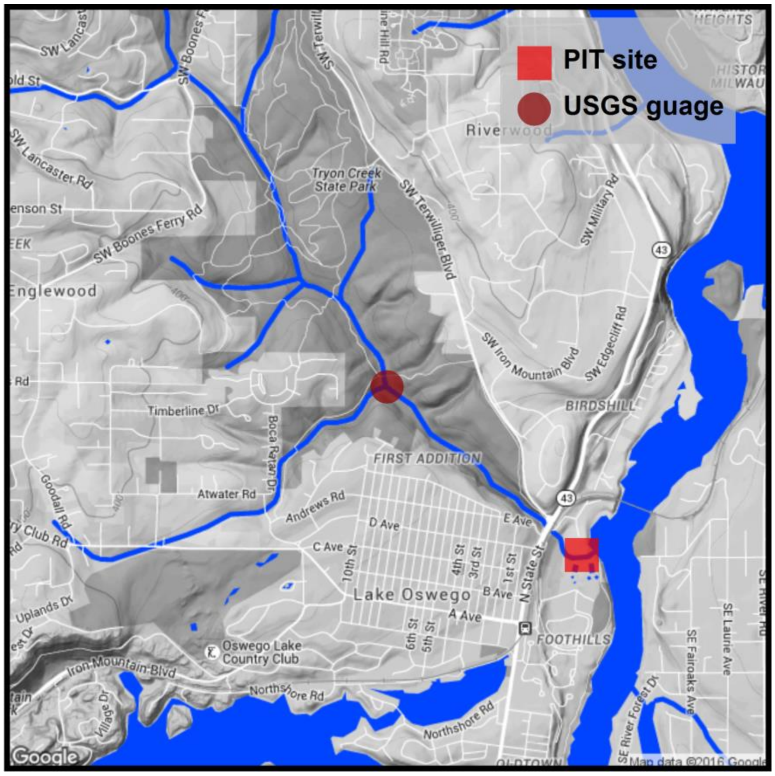

Figure 1: Map of Tryon Creek and locations of PIT antennas and USGS flow gauge.

Methods

Biological Data

In 2015, monthly tagging efforts occurred via electro-fisher and beach seine in Tryon Creek through April and weekly sampling occurred from mid-April to mid-June. Electrofishing consisted of a single pass starting at the Tryon Creek mouth and ending downstream of the Highway 43 culvert pool. Two passes of a beach seine was used to sample the pool. Fish captured using electrofishing were tagged and released 50 meters upstream of Antenna #1 and fish captured using the seine were released back into the pool (Silver et al. 2014).

Biological data recorded during the tagging process were linked to unique PIT tags, uploaded to PTAGIS and stored in a local MS Access database. Biological data included species, fork length (mm), weight (g) and comments (e.g., health, fin erosion, fin clips, etc.).

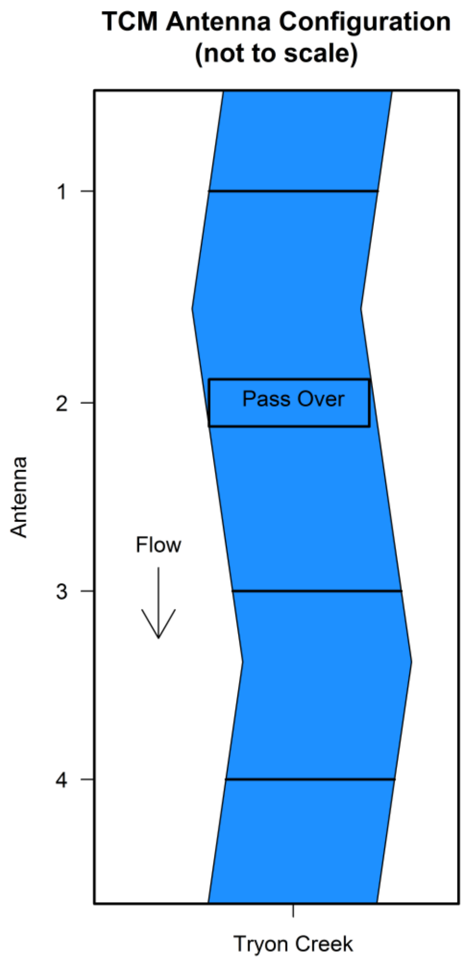

Figure 2: TCM is an in-stream detection system consisting of four PIT anteannas about 25 meters apart located near the mouth of Tryon Creek in Lake Oswego. Each antenna spans the creek width during normal flows. All antennas are pass-through oriented except for Antenna #2 which is a pass over.

Depth

Water depth and temperature were recorded within the study area using HOBO U20 Water Level Data Loggers (Onset). Individual loggers are housed within perforated PVC pipe and suspended approximately 7cm above the substrate. A single logger was installed 1m downstream from Antenna #4, an area influenced by Willamette River backwater. A second logger was installed 1m upstream from Antenna #1, an area rarely affected by backwater. A third reference logger was deployed on land approximately 7m from Tryon Creek to log atmospheric pressure. Water level loggers deployed in stream record water temperature (°C) and absolute water pressure (kPa) at one hour intervals. HOBOware software was used to convert water pressure data to water surface elevation (m) using barometric pressure data collected from the reference logger, and water depth measurements taken adjacent to the logger prior to download.

Detection Coverage

Channel cross sections were modeled to aid our understanding of antenna effective read range given water depth. These models in tandem with depth estimates allow us to confidently approximate percent detection coverage every hour. Cross section data were obtained by staking and leveling a taut string and measuring tape between two points parallel to each antenna at bankfull width (~10 feet high). Height measurements were taken with a stadia rod in foot increments starting at river right. These data were entered into R (R Core Team 2015) and a function was fit to the cross section data points for each antenna (Figure 2). Channel surface area estimates below water levels are calculated by taking the integral of a water level function (f(x)) minus a channel cross section function (g(x)), on the interval between the function intersections (a and b):

$$\int_a^b f(x) - g(x)\, dx$$

The percent detection coverage can be expressed as the read range surface area (antenna plus extended read range) divided by the channel surface area for everything below a given water depth measurement (ft^2).

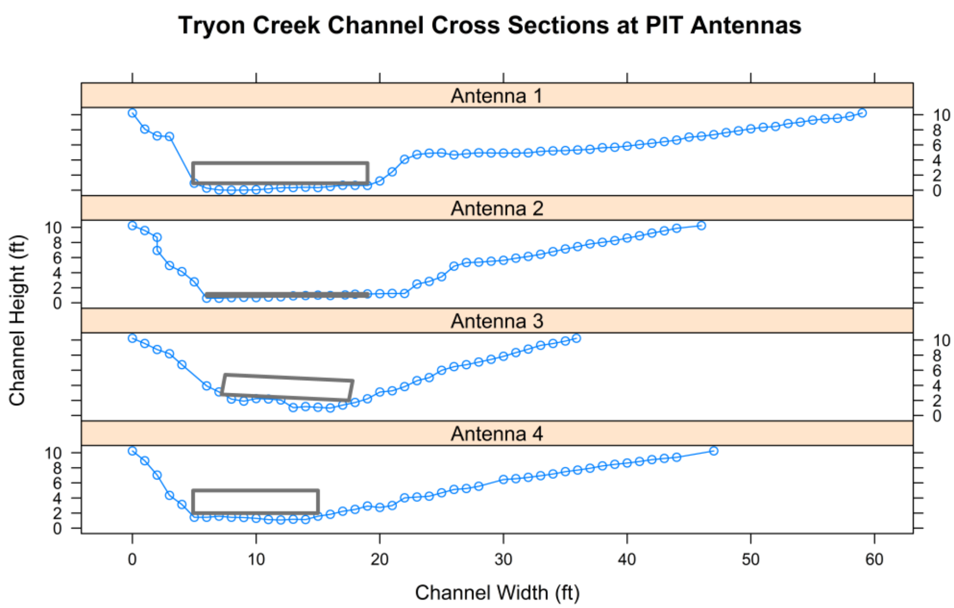

Figure 3: Graphical representation of channel cross sections at each antenna. With depth data, these functions can be used to calculate wetted surface area every hour.

Discharge

Tryon Creek discharge data were obtained via the United States Geological Survey (USGS) water data website (change dates in the URL for a different time period). This gauge is located roughly 1.5 kilometers upstream from the PIT arrays and is maintained by USGS.

Data were also collected from a USGS gauge on the Willamette River approximately 11.5 RKMs downstream from the mouth of Tryon Creek. These data were collected with the intent to characterize how the Willamette influences Tryon Creek, tidally and otherwise. Including discharge, the USGS Willamette gauge records a multitude of water quality measures (i.e., temperature, pH, turbidity, chlorophyll, dissolved oxygen, etc.); although most of these measurements are of little importance to this study, they were collected as some may prove anecdotally valuable.

MUX Diagnostic Reports

Electronic data collected from MUX diagnostic reports included current (amps), relative phase and noise (%) for every antenna as well as data relating to the MUX itself (temperature, tag count, memory, etc.). These data were stored in a buffer every hour. A field laptop connected to the MUX via MiniMon software (www.ptagis.org) also saved and stored buffer data every 24 hours as daily log files. This process automatically converted diagnostic data into a "human readable" format. Having these data stored in two locations has proven helpful as technical mishaps are unavoidable. To deal with the different output formats, a script was written in R that transposes daily log files to match the buffer output (column/row format).

Data Management

- All detection and tagging data are uploaded to PTAGIS and stored locally in their raw form.
- An Access database is in place that stays current by automatically reading in raw detection files upon every open.
- Recapture data is not uploaded to PTAGIS, but can be found in the Access database.
- For a quick view of Tryon Creek detections by species of choice, a web application w/ interactive graph can accessed here: Tryon-Graph.
- Folders containing detection files, Access database, diagnostic files, flow data, water level data and R scripts are located in a single directory on the common drive: "M:\MDTgeneral\PROJECTS\PITdetectionEfficiencyStudy\Tryon data".

Table 1: Calendar of 2015 events in Tryon Creek

Date	Event
12/30/2014	MUX settings were not configured for optimal performance ("unique" not on), which may have resulted in loss of data through 01/06/2015
04/09/2015	Installed HOBO level loggers to record hourly depth measurements
05/14/2015	Took cross channel measurements at Antennas #3 and #4 for detection coverage data (see below)
06/16/2015	Two additional antennas (1 and 2) were installed upstream of the existing antennas (for a total of 4)
06/24/2015	Last tagging event occurred in pool downstream of Highway 43 culvert
06/26/2015	Took cross channel measurements at Antennas #1 and #2
07//07/2015	Beaver constructed a dam between Antennas #1 and #2
12/07/2015	Flows exceeding 600 cfs destroyed everything except Antenna #2 which was configured as a "pass-over"

Analysis

To assess the relative importance of measured variables, the efficiency of Antenna #3 was evaluated by using detections known to be tagged upstream of Antenna #3 and also detected downstream at Antenna #4. Ideally we would have included detections from Antenna #1 and Antenna #2, but they were installed shortly before 2015 tagging efforts halted, resulting in a low number of detections. Detections and non-detections were coded as ones and zeros respectively, and variable data were linked to each observation based on time of detection. If there was more than one variable datum recorded within the duration a tag was detected, the mean of that variable was used.

Several classification models were fit to the data including logistic regression (Logit), k-nearest neighbors (KNN), random forests and boosted trees. The performance of these models was assessed by evaluating mean prediction success rates via k-fold cross-validation. This procedure was used during the variable selection process for every model, as well as to compare the performance of the classification models after variables were fit. The cross validation method was chosen over other methods such as Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC), because it can easily be applied to a diverse set of models, it directly estimates the test error (or success) and makes fewer assumptions about the underlying model (James et al. 2013).

K-fold cross-validation was performed by randomly dividing the 160 observations into 10 sections or “folds” (16 observations per fold). The first fold was used as a validation set and the remaining folds were pooled and used to construct the classification models. The models then predicted the validation set’s response based on its associated predictor variables. This procedure was repeated *k* times until every fold was used as a validation set. The models’ success rate was recorded every iteration and averaged for the final cross validation success rate estimate. K-fold cross-validation was performed using k = 10 because it has been shown empirically that when k = 5 or k = 10, the resulting success rates are not burdened by excessively high bias or variance (James et al. 2013). Other models were fit and evaluated (e.g., linear discriminant analysis and classification trees), but were not pursued further due to either poor performance or near identical success rates and variance.

The relative influence of variables was investigated using three different methods via the logit, random forests and boosted trees models. The logit regression model determined relative importance by summing Akaike weights over all possible models, random forests used the Gini index as a measure of node purity (James et al. 2013), and boosted trees measured the number of times a variable is selected for splitting, weighted by the model's improvement, averaged over all decision trees (Elith et al. 2008).

Results

Detection Summary

There were 16 tagging events that occurred upstream of the Tryon Creek PIT arrays from 31 December 2014 to 24 June 2015 totaling 469 PIT tagged fishes including coho, Chinook, steelhead/trout (*O. mykiss*), and cutthroat trout. Two hundred and nineteen of these tags (46.7%) were detected at the downstream arrays with the bulk of them (72.6%) being from coho (see Table 2). This includes only fish that were tagged and released upstream of the PIT arrays and downstream of the Highway 43 culvert. Tagging events also occurred upstream of the Highway 43 culvert, but these tags were not included in the analysis due to very few detections, likely because the majority of these tagged fished were cutthroat trout and are not expected to emigrate until the following spring.

Table 2: Summary of 2015 PIT tagging data by species, whether they were detected or not, and capture method (E-fisher or seine). There were a few hatchery origin salmon that are lumped together with their wild cohorts and fish identified as hybrids in the field are reported as steelhead in this table.

	Not Detected		Detected		
Species	E-fisher	Seine	E-fisher	Seine	total
Cutthroat	4	3	4	7	18
Chinook	5	12	11	24	52
Coho	36	185	23	136	380
Steelhead	4	0	7	7	18
Whitefish	0	1	0	0	1
total	49	201	45	174	469

There was some uncertainty noted on the data sheets concerning identification of 21 juvenile coho and Chinook salmon, so for the analysis both species were lumped together as a "salmon" group. For a more detailed account of detection timing please visit the interactive graph mentioned in the Data Management section of this document.

Model Parameters

Of all the factors tested, the combination that yielded the highest cross validation success rate for the logistic regression model was discharge, release date (day the fish was tagged) and observed day (day the fish passed arrays). In contrast, all factors were included in the KNN and tree based models, as refining them did not significantly improve or worsen success rates. The KNN model predicts outcomes based on observations nearest to it using a tuning parameter (k) to determine how many surrounding observations to consider. The highest success rate for KNN was found when $k = 9$ (odd numbers are usually selected for k to break ties). The random forests model bootstraps data and creates multiple decision trees, averaging predictions for the final model. In an effort to decorrelate bootstrapped decision trees, the number of potential predictors used at every split in a tree is determined with the parameter m. The best results for the random forests model were achieved when $m = 2$, so that two potential predictors were chosen at random for every split in every bootstrapped decision tree. The boosted trees model fit 5000 trees and used a shrinkage parameter (algorithm learning rate) of $1x10^{-3}$.

Variable Importance

Of the models evaluated, they can accurately predict whether or not an individual tag will be detected at a rate of around 87%, given we know when that tag is passing an antenna (see Table 3). The logistic regression model generally had the highest successful prediction rates and the k-nearest neighbors mean prediction rates had the least variation. Of all the variables measured, discharge was deemed the most important for all three variable importance methods, on average accounting for over 50% of the models' efficacy (see Figure 4).The log odds coefficient for discharge in the linear model was -0.0207 (p-value = 0.002), meaning we would expect the odds of a tag being detected to decrease by 2.1% ($(1 - e^{-0.0207}) * 100$) for every unit increase in flow (cfs) given the release and observed dates are held at fixed values. Time of detection and tagging date were also credited some importance suggesting a possible temporal effect on detection probability. As the date/time of a fish passing the antennas increased, the odds of being detected increased by one percent.

Table 3: 10-fold cross-validation results. Mean success rates and associated variation for four models.

Model	Mean Success Rate	Standard Error
Logit	0.886	0.053
K-Nearest Neighbors	0.867	0.030
Boosted Trees	0.863	0.061
Random forests	0.843	0.057

The variables noise and species had little or no contribution to the models evaluated. Although noise values were considered in the analysis, they were acquired in a sub-optimal way which is likely why they were not considered important to the model. There is a setting on the multiplexer that records noise averages, but the multiplexor was set to record hourly values that were merely snapshots in time. This means we were likely not capturing spikes in conducted and environmental noise. Interestingly, capacitance and phase values were positive contributors to the model, most likely a function of the multiplexer responding to noise values we did not capture. Depth and detection coverage metrics were not considered in the analysis because, a) the depth loggers were launched after most detection had occurred, and b) water levels did not exceed antenna height during the short period of their operation.

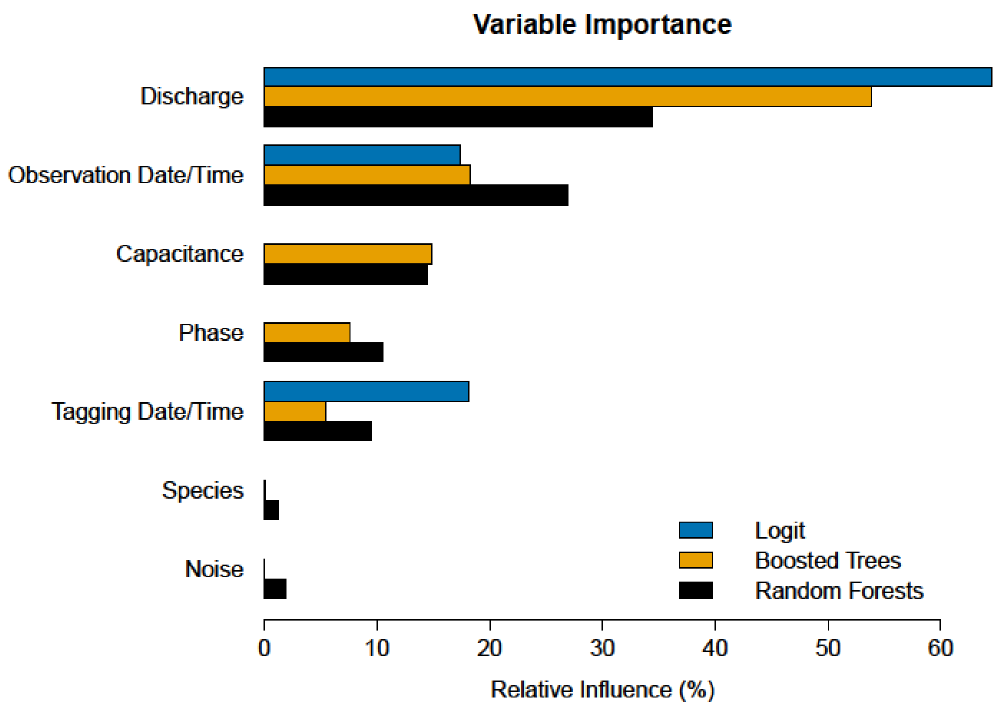

Figure 4: Relative influence of measured variables obtained from the logit, random forests and boosted trees model.

Prediction

To assess whether predicting detection probabilities via explanatory variable information is a viable option, the k-nearest neighbors model was used to generate a binary response (success or failure) based on a more temporally complete data set (not just when fish are known to pass an antenna). The predicted successes and failures were then binned by month and the proportions of successes are reported as efficiency predictions. The k-nearest neighbors model was chosen because it harbored a high success rate and achieved the lowest variance of all the cross validation results. The explanatory variables used to construct the model are flow, capacitance, noise and phase. Variables that would not normally be available with only one antenna present were omitted from the model (e.g., time period a fish passed, but was not detected). Cross-validation success rates did not deviate significantly for the KNN model when these variables were removed.

The predictive model does a decent job of tracking detection probability trends, matching fairly well with the detection history estimates (mean squared error = 0.040) (see Figure 5). Dividing the variables into monthly bins was done arbitrarily; if we intend to use this model to predict future detection probabilities it would make sense to adjust the time scale that it both minimizes the mean squared error and maximizes the precision of detection history estimates.

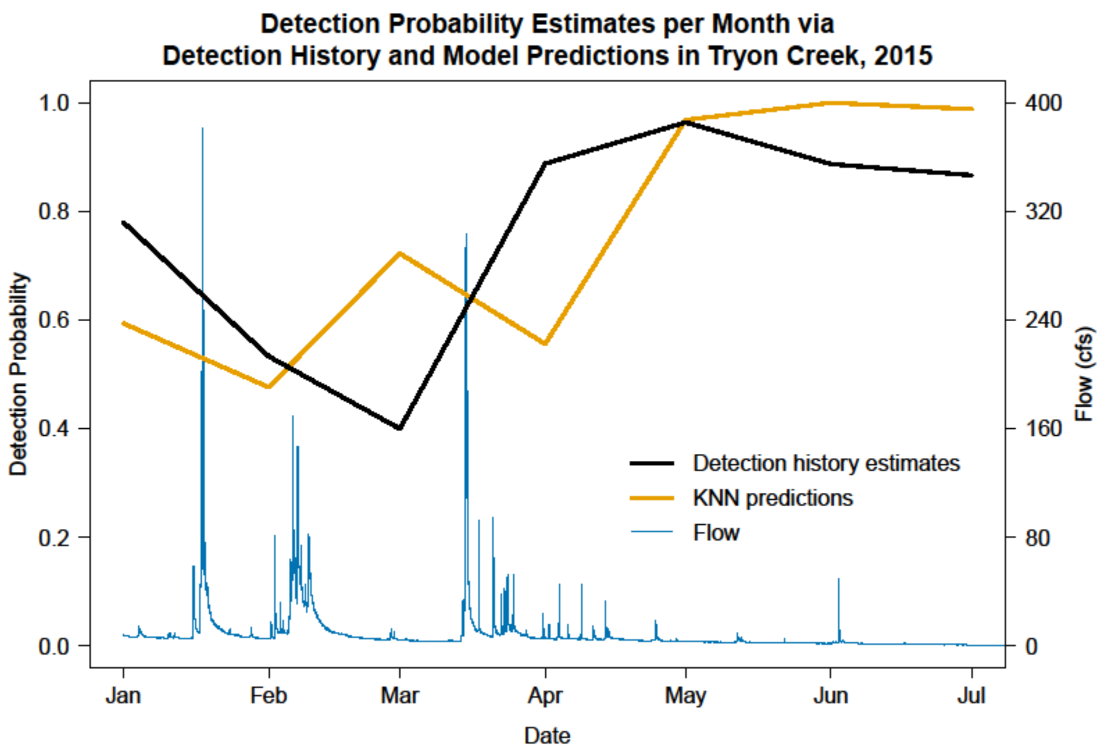

Figure 5: Comparison of monthly detection probability estimates: detection history method (black line) and k-nearest neighbors model (red line).

Discussion

Due to the extremely dry year in 2015, many of the measured explanatory variables were fairly static (e.g., flow and depth). For most of our monitoring, Tryon Creek's water level has not exceeded the height of our arrays, making it difficult to observe any depth associated effects on detection probability. When dynamic conditions were present, tagging efforts had long since halted and detections were minimal. That said, we were still able to capture a relationship between flow and detection probability using data obtained from only two antennas (see Figure

5). While this is not an unexpected relationship, it is a reminder that detection probability is not a fixed parameter, rather a dynamic variable and should be treated as such. Antenna efficiency is often calculated annually which could lead to erroneous estimates when being applied to seasonal migratory movements. The results of these data suggest that estimating detection probability seasonally, taking into account flow, may result in more accurate estimates.

The reason we chose to install one of the antennas configured as a pass-over was to better understand the cost and benefits of pass-over vs. pass-through antennas by comparing operation times and detection probabilities between the two configurations. Pass-over antennas are generally less prone to failure during high flow events due to their low profile, but are limited by decreased read range. Although we were not able to compare the two configurations due to a small sample size, it is interesting to note that on 07 December 2015 Tryon Creek experienced record flows peaking over 600 cfs; all antennas were either broken or dislodged with the exception of the pass-over antenna which not only survived structurally, but continued to run as a functioning detector. The transceiver used to power the detectors and extract data is the Destron Fearing FS1001M, which is the old model slowly being replaced with improved technology such as the IS1001 MTS. Newer technological advances have resulted in an increased antenna read range (often over two feet for large antennas) making the decision to build a pass-over array an easy one, especially in streams prone to high flows.

Based on the prediction results, it may be feasible to use flow as well as antenna diagnostic data in Tryon Creek to estimate antenna efficiency annual trends in cases where only one antenna is present. It would be interesting to test how well a given model performs when applied to different years and/or water sheds. A big caveat to all this information is that the detection history efficiency estimates were obtained using only two antennas. If all four antennas were used in the analysis we would have much more confidence in the detection history estimates (the method used to train and validate the models). There were likely occurrences of fish passing both arrays without being detected, and of course these instances were not taken into account in 2015. Assuming we can maintain four operational arrays during Tryon Creek tagging events in 2016 we would be able to quantify the probability of a tag passing all four antennas without being detected. We would expect this estimate to be minimal which would increase our confidence in efficiency estimates obtained via detection history. Due to problems previously mentioned we are not able to confidently say which antenna performed best (we only calculated efficiency for Antenna #3) and we could not test the importance for all variables we intended (e.g., detection coverage).

Recommendations for 2016

- Ensure all arrays are operational during the 2016 tagging efforts.
- Re-measure channel cross sections at each antenna prior to the 2016 tagging efforts.
- Identify other USFWS PIT interrogation sites monitoring different species (e.g., bull trout in Walla Walla, lamprey in Umatilla) and/or with different site configurations (e.g., Warm Springs LAPS, Clackamas half duplex, etc.) and apply the same methods for comparison.
- Expand analysis, using data from all four Tryon Creek antennas.
- Include depth and detection cover metrics during analyses.

Table 4: Schedule of Activities (2016)

Tasks	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Install/Repair arrays		X	X									
Habitat modeling				X				X				
Record metrics				X	X	X	X	X	X	X	X	X
Watershed comparison				X	X	X						
Analyze data									X	X		
Write progress report											X	X

Acknowledgements

Thanks to Brook Silver (for project collaboration and data support), Ryan Koch, CRFPO staff, everyone who helped out PIT tagging in Tryon Creek, and Alder Creek Kayak and Canoe (for being neighborly and accommodating).

References

Elith, J., J.R. , Leathwick, and T. Hastie. 2008. A working guide to boosted regression trees. Journal of Animal Ecology, *77*(4), pp.802-813.

James, G., D. Witten, T. Hastie, and R. Tibshirani, 2013. An introduction to statistical learning (p. 6). New York: springer.

McCann, J., B. Chockley, H. Schaller, S. Haeseker, R. Lessard, C. Petrosky, E. Tinus, E. Van Dyke, and R. Ehlke. 2015. Comparative Survival Study of PIT-tagged Spring/Summer/Fall Chinook, Summer Steelhead, and Sockeye- 2015 Annual Report. BPA Contract No. 19960200. Available at: http://www.fpc.org/documents/CSS/CSS_2105AnnualReport.pdf

Plumb, J.M., W.P. Connor, K.F. Tiffan, C.M. Moffitt, R.W. Perry, and N.S. Adams. 2012. Estimating and predicting collection probability of fish at dams using multistate modeling. Transactions of the American Fisheries Society, 141(5):1364-1373.

R Core Team. 2015. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Silver, B.P., J.M. Hudson, G.S. Silver, J. Jolley, and T.A. Whitesel. 2014. Tryon Creek Restoration Monitoring, 2013 Progress Report. U.S. Fish and Wildlife Service, Columbia River Fisheries Program Office, Vancouver, W.

U.S. Fish and Wildlife Service
Columbia River Fisheries Program Office
1211 SE Cardinal Court, Suite 100
Vancouver, WA 98683

April 2016
www.fws.gov/columbiariver